

PETE'S THESIS

by

Peter Thompson

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Physics
University of Toronto

Copyright © 2011 by Peter Thompson

Abstract

Pete's Thesis

Peter Thompson

Doctor of Philosophy

Graduate Department of Physics

University of Toronto

2011

(At most 150 words for M.Sc. or 350 words for Ph.D.)

Contents

1	Inclusive Jet Cross Section	1
1.1	Introduction	1
1.2	Jets in ATLAS	3
1.2.1	Topological Clustering	3
1.2.2	Jet-Finding Algorithms	4
1.2.3	Jet Energy Scale Calibration	6
1.2.4	JES uncertainty	9
1.2.5	Jet Selection	12
1.3	Event Selection and Data quality	15
1.3.1	Triggers used for the inclusive jet analysis	15
1.3.2	Transition Bin	18
1.3.3	Dijet Triggers	20
1.4	Data Correction	23
1.4.1	IDS method	24
1.4.2	Matching Efficiency	25
1.5	Treatment of uncertainties	27
1.5.1	Statistical uncertainties	27
1.5.2	Systematic Uncertainties	27
1.6	Results and discussion	29
1.6.1	Inclusive Jet Cross Section	29

1.6.2	Dijet Mass Spectrum	30
	Bibliography	39

Chapter 1

Inclusive Jet Cross Section

1.1 Introduction

The first measurements of the inclusive jet cross section was made in 2010 [5, 6] using 17nb^{-1} of early data. The results described the cross section for production of jets with p_{T} from 60 to 600 GeV and with rapidities less than 2.8 in magnitude. This measurement was repeated in early 2011 using the full 2010 ATLAS dataset (45pb^{-1}) to extend those results, covering jets with p_{T} from 20 GeV to 1.2 TeV. The rapidity range has also been extended into the forward region, to include jets with $|y| < 4.4$. Figure ?? shows the kinematic region covered by the initial measurement compared to that described in this analysis. The dijet mass spectrum has also been measured using 2010 data, and this rapidity coverage of this measurement has been expanded to include the forward region as well.

Section 1.2 describes how jets are defined and calibrated in ATLAS data. The event selection and triggers used for data collection are described in section 1.3, while section 1.4 describes the method used to correct detector resolution effects in the measured data. The treatment of experimental uncertainties is described in section 1.5, and results are presented in section 1.6.

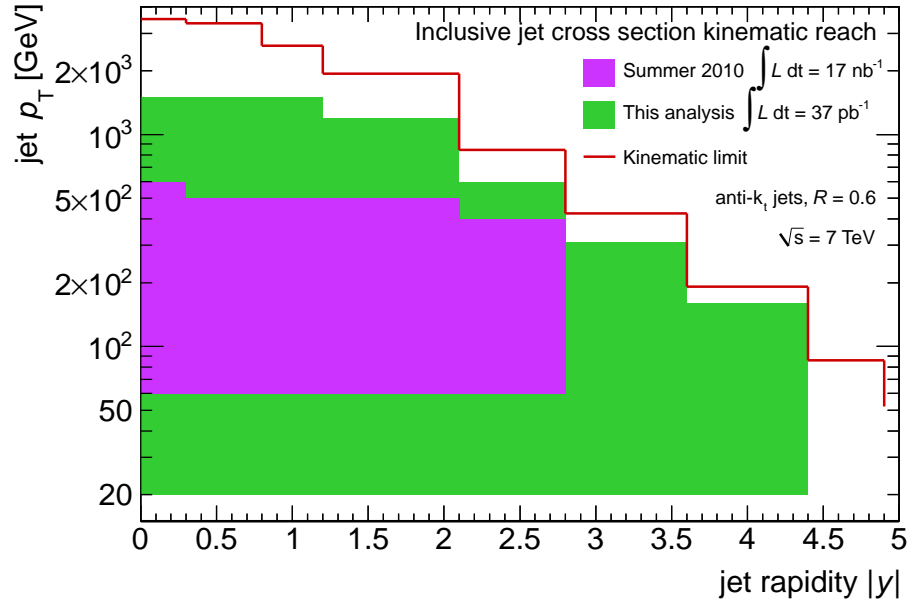


Figure 1.1: Diagram showing the kinematic range covered by the first inclusive jet cross section measurement using 17nb^{-1} (purple) of early data, and that covered when using the full 37pb^{-1} from the 2010 dataset.

1.2 Jets in ATLAS

For the inclusive jet cross section and dijet analyses, jets were reconstructed using the anti- k_t algorithm running on topologically clustered calorimeter cells. The topological and jet clustering algorithms will be described, and then the jet calibration method and the uncertainty on the jet energy scale will be discussed.

1.2.1 Topological Clustering

Topological clusters (or topoclusters) are formed by grouping neighbouring calorimeter cells based on their signal to noise ratio [8]. Cluster seeds are found by searching for calorimeter cells that have an energy greater than some multiple, t_{seed} , of their noise RMS. The noise value used is obtained by adding in quadrature the contributions from electronics noise and pile-up.

Neighbour Cells are added to the cluster provided they are adjacent to the seed cell (or a neighbour cell) and that their signal to noise ratio exceeds the neighbour threshold, $t_{neighbour}$. This is done repeatedly until all neighbour cells are found. Finally, boundary or perimeter cells are added to the cluster by taking all cells that are adjacent to neighbour or seed cells and that have a significance greater than t_{cell} .

Hadronic clusters typically use a “420” scheme, where t_{seed} , $t_{neighbour}$ and t_{cell} have values of 4, 2 and 0, respectively. In this case, the signal is defined as the absolute value of the energy deposited in the calorimeter cell when searching for seed and neighbour cells. This ensures that the contribution from noise is handled symmetrically. A high seed threshold makes it unlikely that a cluster will be seeded purely from noise, while a low cell threshold means that low energy cells around the periphery of the shower are still clustered. A “633” scheme is also used in ATLAS to cluster electromagnetic objects, while other schemes have been investigated using testbeam data (Louise’s thesis). Topoclusters created using the 420 scheme are used as inputs for the jet algorithms.

considered in this analysis.

1.2.2 Jet-Finding Algorithms

The jets considered in this analysis are found using the anti- k_t algorithm [4], with topological clusters used as constituents. The family of k_t -like jet algorithms operate by forming a list of all the constituents in the event. For all constituents and pairs of constituents, the jet resolution quantities

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \quad (1.1)$$

and

$$d_i = k_{ti}^{2p} \quad (1.2)$$

are computed, where k_{ti} , ϕ_i and y_i are the transverse momentum, azimuthal angle and rapidity of the i th constituent, respectively. The distance parameter R describes the desired size of the jets being sought after, such that found jets will be separated by no less than R in (y, ϕ) -space. The parameter p defines the jet-finding algorithm: $p = 1$ corresponds to the k_t algorithm, $p = -1$ gives the anti- k_t algorithm, and $p = 0$ corresponds to the Cambridge-Aachen cone algorithm.

Once d_{ij} and d_i have been computed, the values are sorted. If the smallest value present corresponds to a d_{ij} , then the i -th and j -th constituents are merged into a proto-jet. This is done by summing their four momenta. The proto-jet is then added to the list of constituents while its components are removed from the list. The d_{ij} and d_i values are then recomputed for all remaining constituents and proto-jets. This process is repeated, with each iteration either adding constituents to an existing proto-jet or merging constituents to form a new proto-jet. In the case where a d_i value is smaller than any d_{ij} value, then this proto-jet is considered a complete, final jet and it is removed from the list. The process continues until the constituent list is empty, with all constituents

having been used to form complete jets.

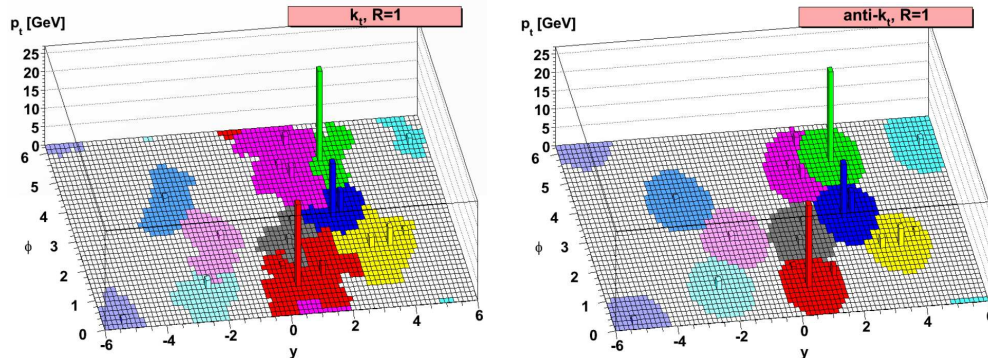


Figure 1.2: jets found by the k_t algorithm (left) and the anti- k_t algorithm (right). The event was generated using HERWIG [?], and contains some soft radiation in addition to the high-pt constituents

In the k_t algorithm d_{ij} is approximately equal to the relative difference in transverse momentum between the two constituents, in the limit where the angle between the constituents is small. The k_t algorithm thus acts to cluster together constituents with similar momenta, reversing the fragmentation process caused by QCD. However, this procedure clusters constituents with the smallest transverse momenta first, such that high p_T constituents may be clustered around groups of low p_T constituents. This tends to result in jets with irregularly shaped boundaries. Conversely, the anti- k_t algorithm considers the highest p_T constituents first and builds the proto-jets around those, resulting in conically-shaped jets (fig. 1.2(b)). This procedure also means that jets found by the anti- k_t algorithm tend to be less sensitive to the effects of pile-up and the underlying event [4].

The k_t -like jet algorithms are favored because they are both infrared (IR) and collinear safe. IR safety means that the jets found are stable with respect to low p_T constituents arising from pile-up or the underlying event (fig ??). For an algorithm to be IR safe, the presence of soft particles can not “confuse” the algorithm into mistaking two separate jets for a single large jet if there is some soft radiation between them (fig. 1.4). Collinear safety requires that the algorithm will still find a jet if, for instance, a single high-pt constituent

is replaced by two (or more) close together constituents with lower p_T (fig 1.3). The k_t -like algorithms have these properties, whereas more iterative cone algorithms tend not to.

1.2.3 Jet Energy Scale Calibration

Once a jet has been defined, its energy must be calibrated to the Jet Energy Scale (JES). This is done through an “EM+JES” scheme, whereby the EM scale energy of the jet is multiplied by a correcting factor to obtain the energy at the JES. The calibration is derived from Monte Carlo using a numerical inversion process. This calibration accounts for the following effects:

- non-compensation of the calorimeter, i.e. the energy is calibrated to the hadronic scale.
- energy deposited in inactive (dead) regions of the detector
- leakage effects from particle showers not fully contained in the calorimeters
- particles contained in the truth jet but not in the reco jet
- energy from showering particles that is not collected by the topoclustering algorithm. (out of cluster corrections).

The calibration is done in three steps. First, the EM scale energy of the jet is adjusted in order to correct for pile-up effects. Additional proton-proton interactions from the same event can deposit energy in the calorimeter, effecting the energy of the high p_T objects from the hard scattering. Minimum bias data is used to determine the average energy deposited in the calorimeter as a function of pseudorapidity and the number of primary vertices reconstructed from the event. This information is used to subtract the average EM scale energy added to the jet as a result of pileup.

The second step is to correct the kinematics of the jet, still at the EM scale, based on the location of the hard scattering. The vertex with the highest sum of squared transverse

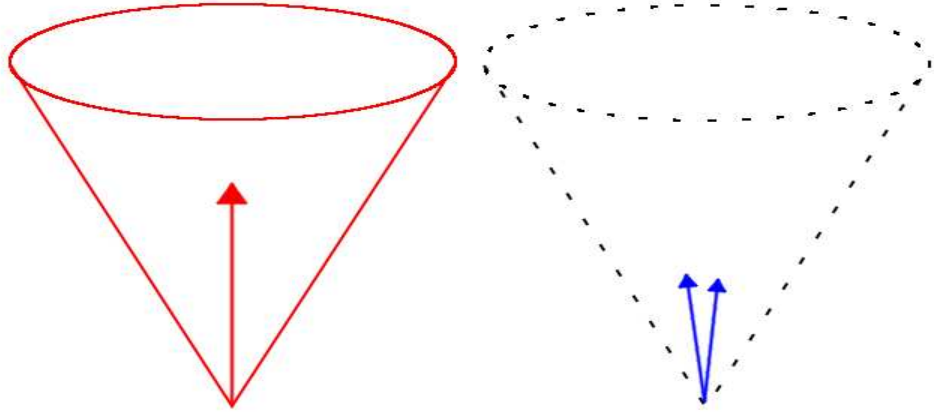


Figure 1.3: An example of an algorithm which is not collinear safe. On the left, a jet is found around a single high-pt constituent. On the right, the single constituent is replaced by two, each with half the p_T of the original. In this case, the algorithm fails to find a jet.

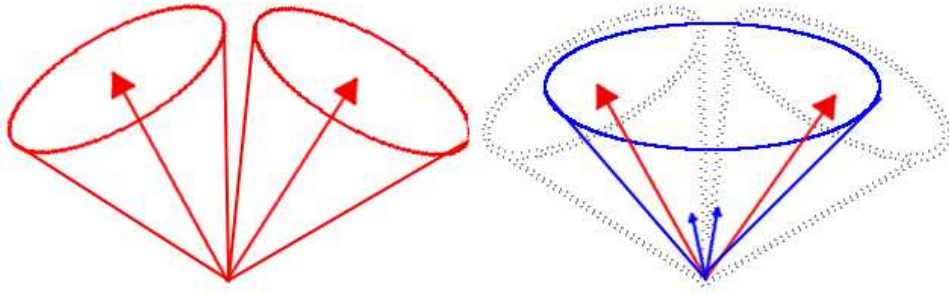


Figure 1.4: example of a jet algorithm which is not IR safe. In the figure on the left, two distinct jets are found around the high-pt constituents. In the presence of soft radiation (right), The algorithm finds only a single jet.

momenta from tracks ($\sum p_{T,track}^2$) is taken as the position of the hard scatter, and all jet kinematic quantities (p_T, y, ϕ , etc) are recomputed using this vertex as the origin.

Finally, the JES correction is applied. The correction is derived exclusively from Monte Carlo, using samples generated without the inclusion of pile-up. An event generator is used to simulate a hard parton-parton scattering typical of proton-proton collisions, which outputs a set of final state particles and their four momenta. Geant 4 is then used to simulate the interaction of these particles with the detector. The event is then reconstructed using the same methods that are used for the data: calorimeter cells are reconstructed and used to form topological clusters, on which jet finding algorithms are run. Truth jets are formed by running the jet finding algorithms on the final state particles output by the event generator.

The calibration is derived by comparing the jets reconstructed from the (simulated) calorimeter information to the truth jets. A reconstructed jet is matched to a truth jet if the distance between them is less than 0.3 in (y, ϕ) . Only isolated jets are matched: a reconstructed/truth jet must have no other jets with $p_{T,EM} > 7\text{GeV}$ within $2.5R$, or it is not used in the derivation of the calibration.

The matched reconstructed-truth jet pairs are then used to define the response R , such that

$$R = \frac{E_{\text{reco}}^{\text{EM}}}{E_{\text{truth}}}, \quad (1.3)$$

where $E_{\text{reco}}^{\text{EM}}$ is the EM scale energy of the reconstructed jet and E_{truth} is the energy of the truth jet. The response is binned in E_{truth} and η_{det} , the pseudorapidity of the reconstructed jet at the EM scale. For each $(E_{\text{truth}}, \eta_{\text{det}})$ bin, the mean reconstructed jet energy, $\langle E_{\text{reco}}^{\text{EM}} \rangle$, is found, and a gaussian fit is used to extract the mean response, $\langle R \rangle$.

The response is then parameterized as a function of $E_{\text{reco}}^{\text{EM}}$ for each η_{det} bin. For the k -th η_{det} bin, a fit is performed on the $(\langle E_{\text{reco}}^{\text{EM}} \rangle, \langle R \rangle)$ points obtained from each E_{truth} bin. The fitted function is of the form

$$F_{\text{calib},k}(E_{\text{reco}}^{\text{EM}}) = \sum_{j=0}^{N_{\text{max}}} a_j (\log E_{\text{reco}}^{\text{EM}})^j, \quad (1.4)$$

where the a_j are free parameters and N_{max} is an integer between 1 and 6 chosen to maximize the goodness of fit.

The correction factor is then found by inverting $F_{\text{calib},k}$, so the final EM+JES energy of a jet lying in the k -th η_{det} bin is given by

$$E_{\text{reco}}^{\text{JES}} = \frac{E_{\text{reco}}^{\text{EM}}}{F_{\text{calib},k}(E_{\text{reco}}^{\text{EM}})}. \quad (1.5)$$

1.2.4 JES uncertainty

The systematic uncertainty on the JES is an important quantity, and one of the dominant sources of uncertainty in this analysis. There are several sources that contribute to this uncertainty, outlined below:

- Uncertainty from the calibration method** The method assumes that all jet constituents require the same average correction. Additionally, the same correction factor is used for both the energy and transverse momenta of the jet, which may bias the calibrated p_T in cases where the calibrated jet mass differs from the mass of the truth jet. The uncertainty arising from the calibration method is estimated by comparing the reconstructed jets at the EM+JES scale to their truth jet counterparts. The responses $\langle RE \rangle = \langle E_{\text{reco}}^{\text{EM+JES}} / E_{\text{truth}} \rangle$ and $\langle RP \rangle = \langle p_{T,\text{reco}}^{\text{EM+JES}} / p_{T,\text{truth}} \rangle$ are computed, and binned in terms of $p_{T,\text{reco}}^{\text{EM+JES}}$. Any deviation of $\langle RE \rangle$ or $\langle RP \rangle$ from unity suggests that the kinematics of the reconstructed jets after calibration are not equal to those of the truth level jets, for the reasons mentioned above. The uncertainty assigned to the calibration method is taken as the largest deviation of $\langle RE \rangle$ or $\langle RP \rangle$ from unity, and is found to be less than 1% for $p_{T,\text{reco}}^{\text{EM+JES}} > 30 \text{ GeV}$.

- **Uncertainty from calorimeter response** The contribution to the JES systematic from the uncertainty in the calorimeter response is derived from single particle measurements. The uncertainty in the response to charged hadrons is measured in E/p studies [1] and in testbeam data [7]. The simulation framework allows the particles in the truth jet to be associated with the energy they deposit in the calorimeter, and thus the single particle response uncertainties can be propagated to obtain an uncertainty for the response of the jet. When estimating this uncertainty, effects relating to the calorimeter acceptance, charged particles with $E > 400\text{GeV}$, and energy deposited by neutral hadrons are also considered. Effects related to the calorimeter response are found to contribute a $1.5\% - 4\%$ uncertainty to the JES systematic.
- **Uncertainty from noise thresholds in detector simulation** The noise present in the calorimeter electronics can change over time, whereas the noise used in the simulation is fixed when the MC sample is generated. The effect of the noise threshold in the simulation was measured by increasing the noise thresholds for the topoclustering algorithm by amounts of 5-10%. This influences which cells are grouped into topoclusters, and thus contribute energy to the jet. The uncertainty assigned to this effect was found to be negligible for jets with $p_T > 45\text{GeV}$, and is estimated as 1-2% for jets with lower p_T .
- **Effect of additional material in simulation** As the JES calibration is intended to correct for the effects of inactive material, it is sensitive to the material description of the detector in the simulation. The effects of this were estimated by adding additional material to the simulation geometry in several places, and comparing the response obtained with the modified geometry to that obtained using the nominal geometry.

- **MC event generators** The nominal MC sample used in the derivation of the JES was generated using PYTHIA, using the AMBT1 tune. Samples were also produced using ALPGEN interfaced with HERWIG and JIMMY, and using the perugia 2010 tune in PYTHIA. The ALPGEN sample used the CTEQ6.1 pdfset, and treated parton showering and hadronization effects differently to the nominal PYTHIA sample, whereas the perugia2010 sample provided a different treatment of the underlying event. Deviations between the response obtained from the perugia2010 and ALPGEN +HERWIG +JIMMY samples with respect to the nominal PYTHIA sample were used to estimate the uncertainty arising from the choice of theoretical physics models.
- **Relative calibration of uncertainties between forward and central regions.** Contributions to the JES uncertainty from the above sources have been calculated in the central region, $0.3 < |\eta| < 0.8$. This uncertainty is used as a baseline, and an intercalibration method [2] is used to extend the estimate of the JES systematic into other pseudorapidity regions. This method uses a p_T balancing technique applied to dijet events in order to obtain the ratio of the calibrated jet responses in different regions of pseudorapidity. This response was calculated for data and simulation, using several different MC event generators. The RMS of the differences in response between MC and data is then added in quadrature to the baseline uncertainty, yielding the uncertainty in higher pseudorapidity regions.

The total JES uncertainty, and its components, are plotted in figures 1.5 and 1.6. The dominant contribution in the central region ($0.3 < |\eta| < 0.8$) is from the uncertainty in the calorimeter response. In other pseudorapidity regions the intercalibration uncertainty dominates at low p_T .

1.2.5 Jet Selection

After Jets have been calibrated, there are some criteria that they need to meet before being included in the analysis. Certain detector issues were capable of causing a jet to be reconstructed even if there were no physical particles depositing energy in that region of the calorimeter. Jet cleaning cuts were made to address these issues, and remove from the analysis as many “fake” jets as possible. The jet cleaning cuts, and the problems they are intended to address, are listed below

- **Coherent noise in the EM calorimeter** Jets with $|\eta| < 2.8$ were rejected if EM_f , the fraction of energy deposited in the EM calorimeter, exceeded 0.9 while the LAr quality variable exceeded 0.8. The LAr quality refers to the fraction of LAr cells in the jet which have a pulse shape significantly different to a reference shape.
- **Noise bursts (“spikes”) in the HEC** Jets were rejected if the fraction of energy deposited in the HEC was greater than $1 - HECQ$, where $HECQ$ is the HEC quality variable and is defined analogously to the LAr quality variable. Jets were also rejected if the sum of negative energy cells exceeded 60 GeV in magnitude.
- **cosmics/beam background** Jets were rejected if the average timing for jet cells was greater than 10ns from the average event time. For jets with $|\eta| < 2.0$, tracking information can be used to define the charged fraction, Chf , which is the fraction of the jet p_T associated with tracks in the inner detector. In this case, Jets were rejected if the charge fraction were less than 0.1 and the EM fraction was less than 0.05, if the EM fraction exceeded 0.95 while the charge fraction was less than 0.05, or if 99% of the jet energy was deposited in a single layer of the calorimeter. Jets with $|\eta| > 2.0$ were rejected if the EM fraction was less than 0.05.

Jets which failed these cleaning cuts were not included in the analysis.

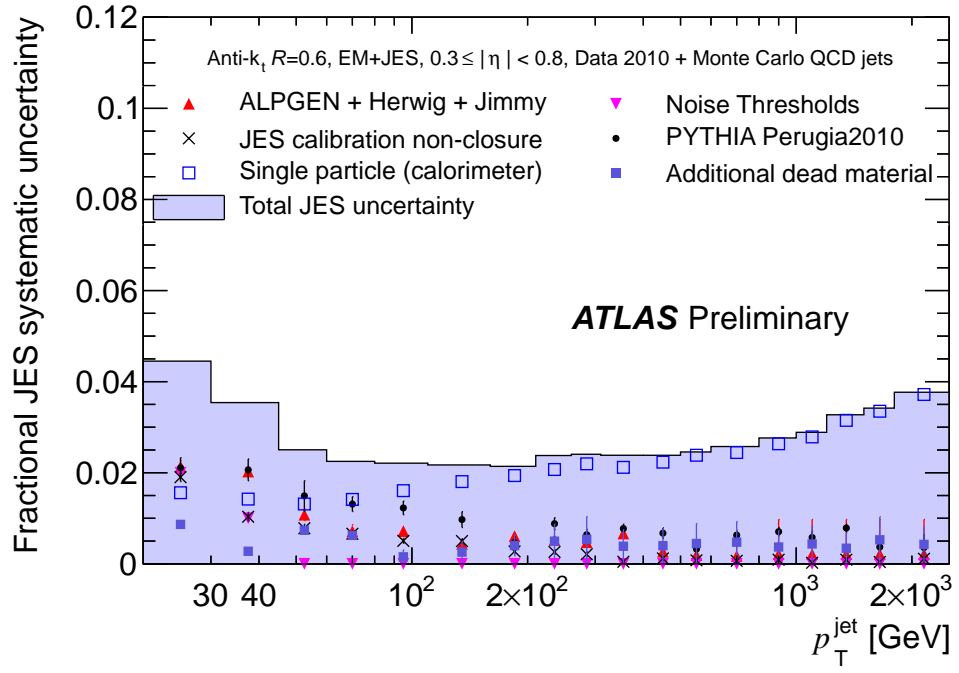


Figure 1.5: JES uncertainty in the central barrel region ($0.3 < |\eta| < 0.8$)

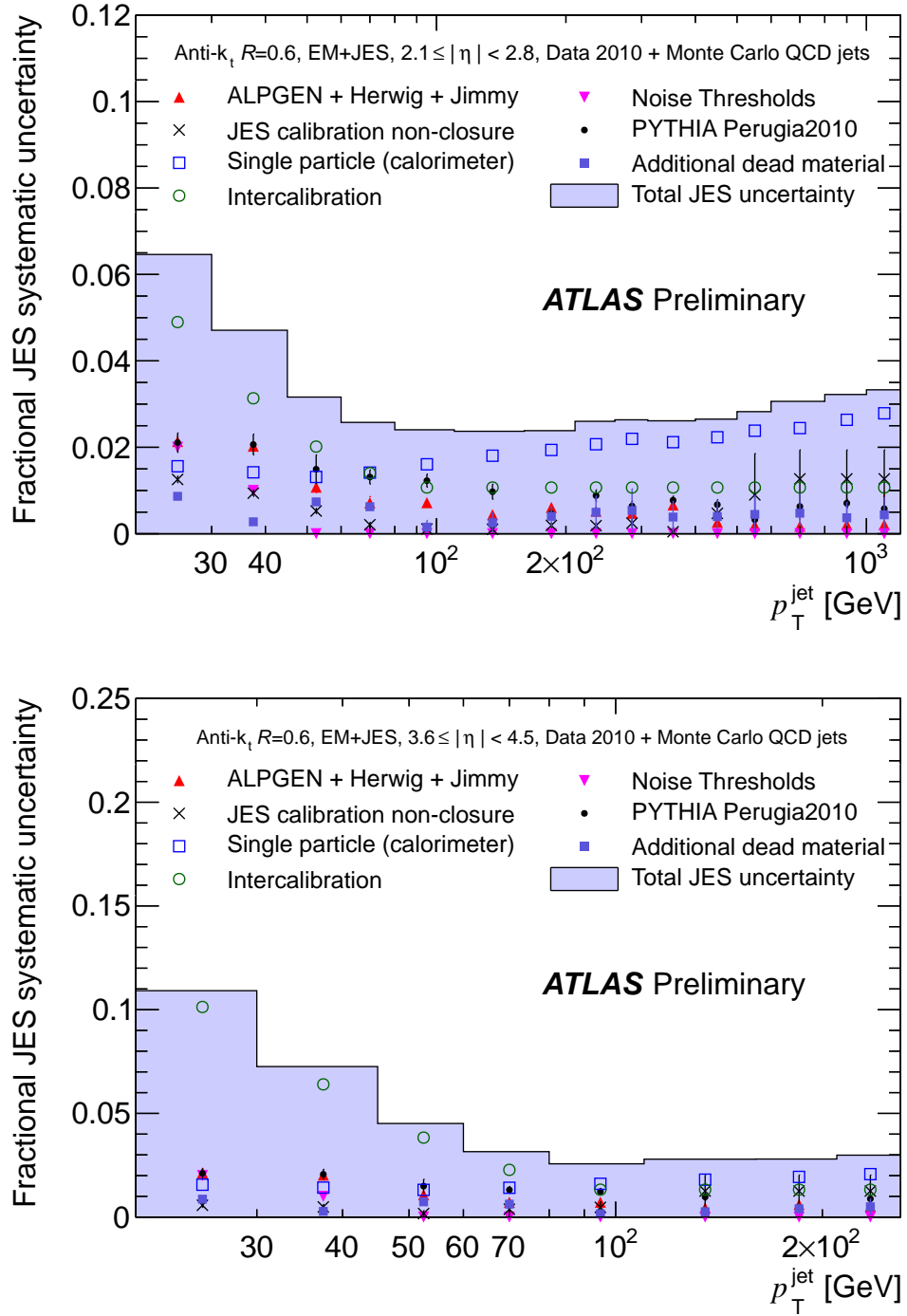


Figure 1.6: JES uncertainty for the end-cap ($2.1 < |\eta| < 2.8$) and FCal ($3.1 < |\eta| < 4.5$). At low p_T , the intercalibration provides the dominant source of uncertainty.

1.3 Event Selection and Data quality

In order for an event to be considered in the inclusive jet cross section analysis, data quality flags for the event needed to be green. The event was also required to have a primary vertex with at least 5 tracks. Each p_T -rapidity bin in the inclusive analysis was associated with a trigger. An event only contributed jets to a given p_T - y bin if the trigger threshold associated with that bin is passed.

The same criteria were used for the dijet mass spectrum analysis, with the addition that leading jet in the event was required to have $p_T > 30$ GeV, while the subleading jet was required to have $p_T > 20$ GeV.

1.3.1 Triggers used for the inclusive jet analysis

The cross section measurement is divided into 7 bins of rapidity and 16 bins of p_T . A dedicated, fully efficient trigger is used to collect jets in each bin in order to maximize statistics. The three lowest p_T bins are filled using data taken from the MBTS1 (minimum bias) trigger. Minimum bias data from only the three earliest periods of running (A-C) was used for this purpose, as in these periods the trigger had a lower prescale (and so collected more data) and there was a minimal amount of pile-up. At higher p_T (above 60 GeV), data is taken using the central jet trigger for jets with $|y| < 2.8$. The rapidity bin from $2.8 < |y| < 3.6$ is referred to as the “transition bin”, as it covers the transition region between the end cap calorimeters and the FCal (figure 1.7). In this region both the central and forward jet triggers are used, whereas in the forward region ($3.6 < |y| < 4.4$) only the forward jet trigger is used. The behaviour of the jet trigger system needs to be understood before they can be used effectively. This is achieved by measuring the trigger efficiency. The inclusive efficiency of a trigger is defined as

$$\epsilon_{inc} = \frac{N_{triggered}}{N_{reference}}, \quad (1.6)$$

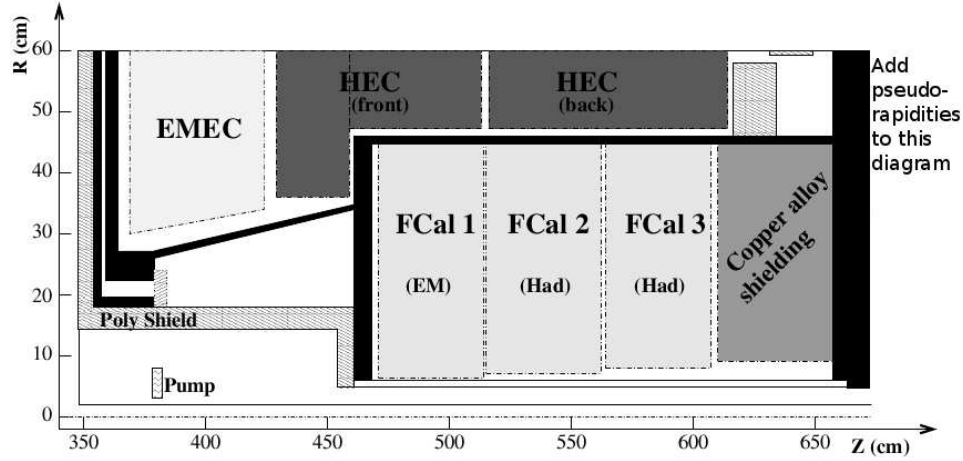


Figure 1.7: diagram showing cross section of part of an end-cap in ATLAS. The forward bin corresponds approximately to the region $2.8 < |\eta| < 3.6$, and is completely contained within the FCal. The transition bin covers the area between $|\eta| = 2.8$ and $|\eta| = 3.6$, which is partially occupied by both the end-cap calorimeters and the FCal.

where $N_{reference}$ is the total number of jets reconstructed offline from a reference sample, and $N_{triggered}$ is the number of reconstructed jets coming from events in the reference sample which pass the trigger condition. This efficiency is then binned in the p_T of the reconstructed jets to produce a turn-on curve. Trigger efficiencies for the transition and forward bins are plotted in figures 1.8 and 1.9, respectively.

The plateau point is defined by integrating bins of $N_{triggered}$ and $N_{reference}$ from high p_T towards low p_T , until reaching the point where the ratio of these sums drops below 99%. The high edge of this bin is then taken as the plateau point, as the trigger is at least 99% efficient at p_T 's higher than this value. Above the plateau point the trigger is assumed to be fully efficient, and is used to collect data for the appropriate p_T bins.

While studying the forward jet triggers, it was observed that events containing high p_T jets in the region $x < y < z$ and $x < \phi < z$ were not being selected by the trigger. The L1 trigger tower in this region was producing very little signal, and was effectively dead. This reduced the geometrical acceptance of the forward jet trigger by $\frac{1}{128}$, which is corrected for in the cross section calculation. Consequently, the plateau points in the

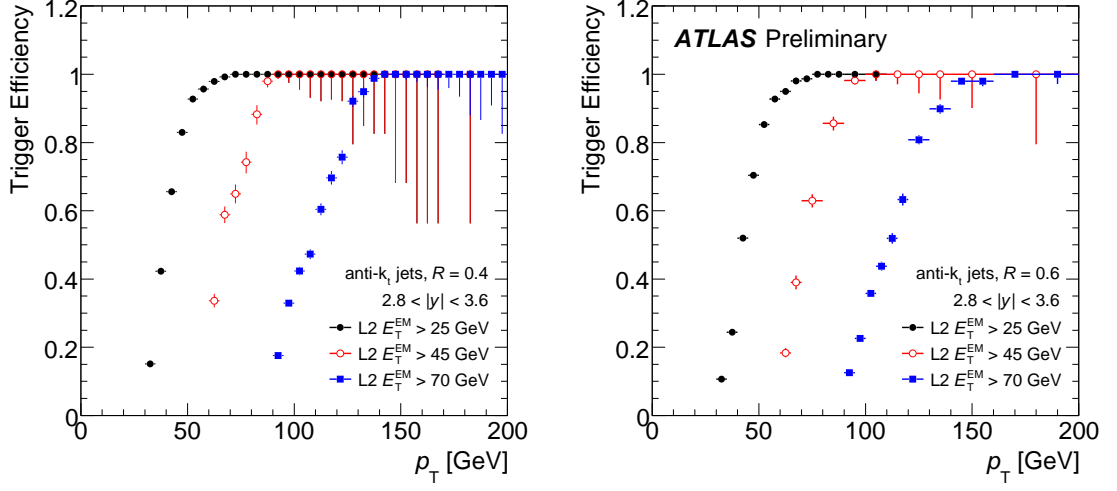


Figure 1.8: Efficiencies for L2 triggers in the transition bin, for anti- k_t jets with $R=0.4$ (left) and $R=0.6$ (right).

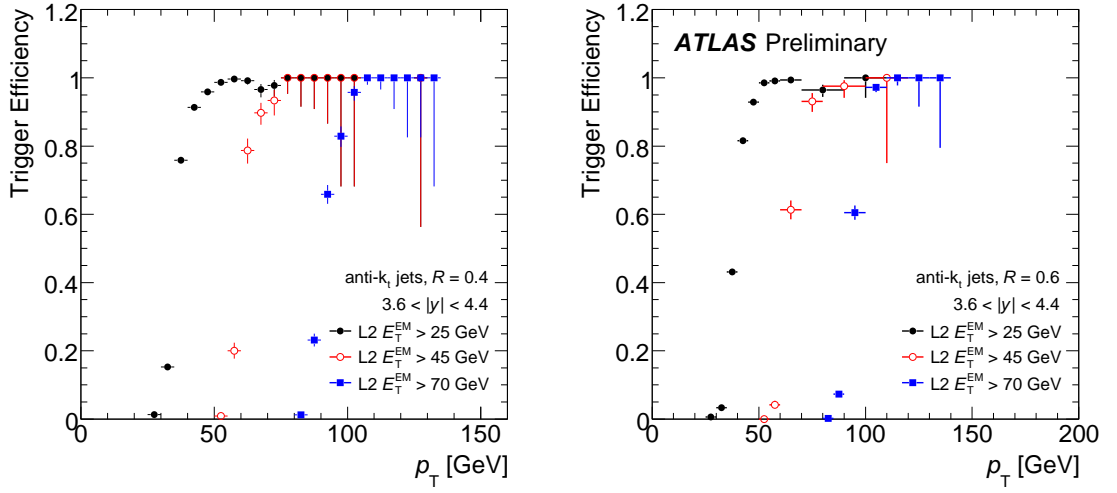


Figure 1.9: Efficiencies for L2 triggers in the forward bin, for anti- k_t jets with $R=0.4$ (left) and $R=0.6$ (right).

forward and transition bins are defined at an efficiency of 98%. It should be noted that while the trigger tower is considered dead, calorimeter cells (and hence jet reconstruction) are unaffected in this region.

As data taking progressed, the trigger system was commissioned further. The forward jet triggers were commissioned after run **1561127** (period E5), and used to collect data in the forward and transition regions from then on. The HLT system was commissioned from period G on, which allowed event rejection to occur at L2. The EF remained in pass-through mode throughout the remainder of 2010, and so all events passing at L2 were recorded. Triggers at L1 were used to collect data until the HLT was commissioned, after which L2 triggers were used.

1.3.2 Transition Bin

The transition bin covers the region between end cap calorimeters and the FCal. Information from the EMEC and HEC is only used by the central jet trigger, while information from the FCal is only used by the forward jet trigger. In order for the analysis to be sensitive to all jets in this region of rapidity, both the central and forward jet triggers need to be considered when collecting data. This is done by taking the logical “OR” of the two triggers, such that jets in the event are counted if either the central or forward trigger condition is met. The efficiency of each trigger through this rapidity range is shown in figure 1.10.

When only a single trigger is used, the measured cross section is given by

$$\sigma = N_{jets} \left(\sum_i \frac{\mathcal{L}_i}{S_i} \right)^{-1}, \quad (1.7)$$

where N_{jets} is the number of jets observed and \mathcal{L}_i and S_i are the integrated luminosity and trigger prescale for the i th luminosity block, respectively. . The situation is more complex in the case where two (or more) triggers are collecting data. In this situation we classify

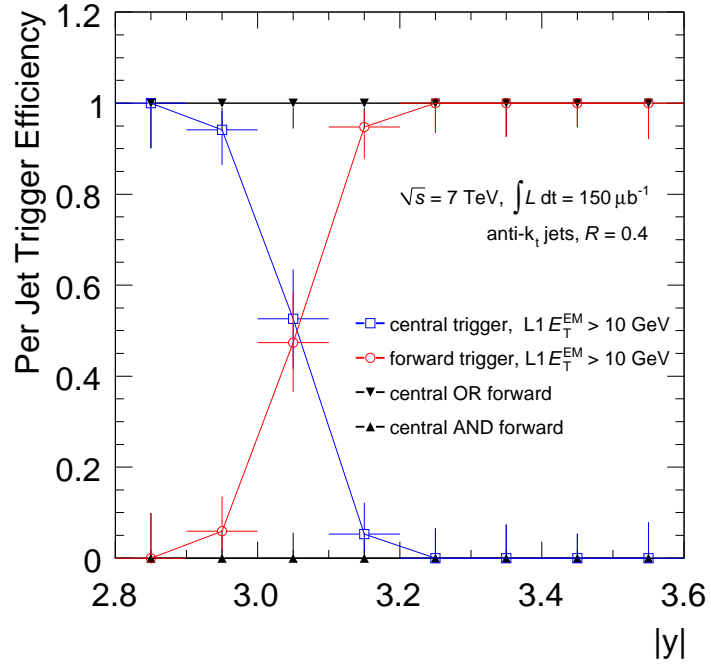


Figure 1.10: Efficiency of Level 1 forward (red) and central (blue) triggers as a function of offline jet rapidity for jets with p_T greater than 45 GeV in the transition bin ($2.8 < |y| < 3.6$). Note that the OR of these remains fully efficient throughout the rapidity range.

events as belonging to one of three classes, based on whether the event passed the forward trigger condition but not the central (case 10), the central trigger condition but not the forward (case 01), or both trigger conditions (case 11). When making these distinctions, it is important to consider whether the event satisfied a trigger condition before the prescale was applied, as events should be classified according to the triggers that they could potentially be accepted by rather than the triggers which actually recorded the event. For L1 triggers the “Trigger Before Prescale” flag can be checked in athena, but for L2 triggers the jets found at L2 must be checked to see if any exceed the trigger threshold.

Once the events have been divided into their respective classes, the cross section may be written as

$$\sigma = \frac{N_{01}}{\mathcal{L}_{01}} + \frac{N_{10}}{\mathcal{L}_{10}} + \frac{N_{11}}{\mathcal{L}_{11}}, \quad (1.8)$$

where integrated luminosities are given by

$$\mathcal{L}_{01} = \sum_i \frac{\mathcal{L}_i}{S_{i,01}} \quad (1.9)$$

$$\mathcal{L}_{10} = \sum_i \frac{\mathcal{L}_i}{S_{i,10}} \quad (1.10)$$

$$\mathcal{L}_{11} = \sum_i \frac{\mathcal{L}_i (S_{i,01} + S_{i,10} - 1)}{S_{i,01} S_{i,10}} \quad (1.11)$$

and $S_{i,10}$ ($S_{i,01}$) is the prescale of the forward (central) jet trigger for the i th lumi-block.

1.3.3 Dijet Triggers

For the initial measurements made using data from ATLAS [5, 6] the dijet mass spectrum was binned in terms of the dijet mass, m_{12} , and the maximum rapidity of the two jets, $y_{max} = \max(|y_1|, |y_2|)$. Dijet events were selected by triggering on the leading jet, using the same trigger thresholds as for inclusive analysis. These first measurements only considered cases where both jets were in the central region, with $y_{max} < 2.8$.

When the dijet analysis was extended to cover the transition and forward regions, a new trigger scheme was considered. The intent was to bin the trigger efficiency in terms of the observables of interest. The central jet trigger would be used for $y_{max} < 2.8$, the forward jet trigger for $3.6 < y_{max} < 4.4$, and the OR of the two for $2.8 < y_{max} < 3.6$, with the trigger efficiency being described by a turn-on curve in m_{12} . Each trigger could then be considered fully efficient above some threshold value for the dijet mass, and used to collect data above this point. This trigger scheme was later abandoned, as the trigger system is based on jet p_T and thus inflates the minimum dijet mass at which the trigger becomes unbiased. Consider the case where both jets in the dijet system have transverse momenta well above the trigger threshold, but the separation between jets is small. The small separation between jets will yield a small mass, and the trigger will efficiently accept dijet events in this configuration. However, in cases where the separation between jets is large and the jet momenta are close to the trigger threshold, the resulting value for the dijet mass can still be quite large while the trigger is not fully efficient.

For this reason, the observables of interest were changed to the dijet mass and the jet separation, $y^* = |y_1 - y_2|/2$. The trigger scheme then reverted back to that used for earlier versions of the analysis wherein the leading jet is used to trigger the event, although this was extended to cover the forward region. The L1 trigger efficiencies for dijet events are plotted as a function of the leading jet p_T in figures 1.11 and 1.12, for the transition bin and forward bin respectively.

It was then expanded to consider the subleading jet, utilizing a method similar to what was done for the transition bin in the inclusive analysis. The kinematic region is divided into a number of bins in p_T and rapidity, each of which is associated with a trigger threshold. Events are then divided into categories based on whether the trigger condition associated with the leading jet was met, or that for the subleading jet, or if both trigger conditions were satisfied. As there is some overlap between the forward and central jet triggers in the region $3.0 < |y| < 3.2$ (fig 1.10), jets in this region are

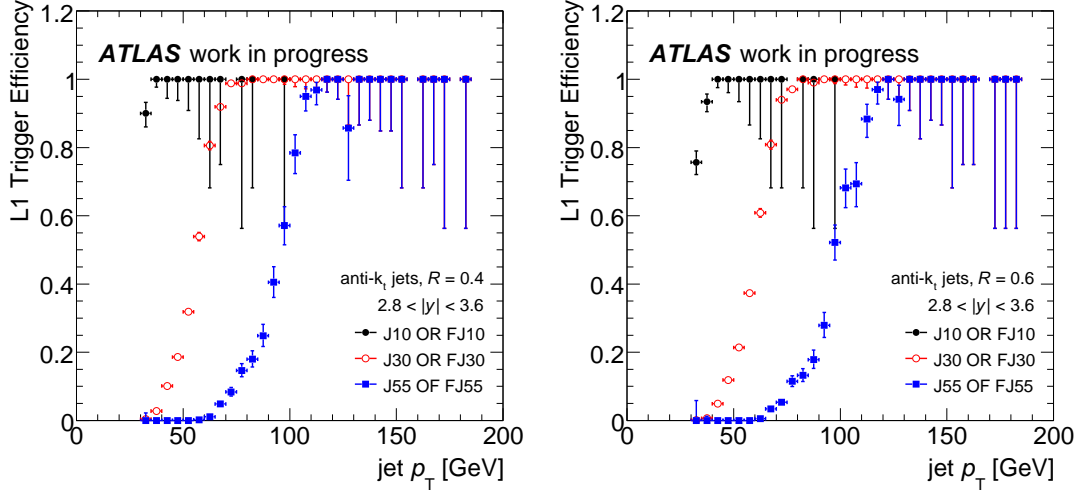


Figure 1.11: L1 trigger efficiencies for dijet events in which the leading jet lies in the transition bin, for anti- k_t jets with $R = 0.4$ (left) and $R = 0.6$ (right).

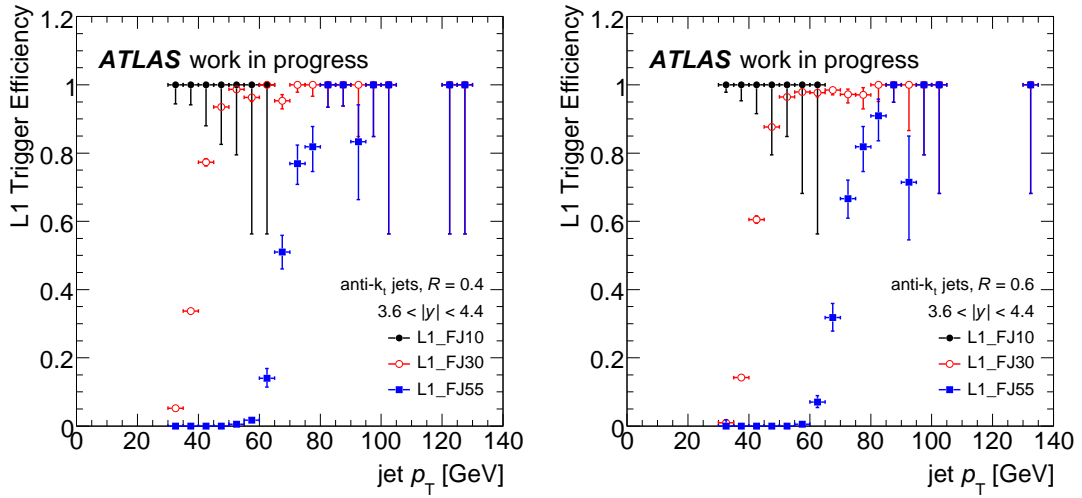


Figure 1.12: L1 trigger efficiencies for dijet events in which the leading jet lies in the forward bin, for anti- k_t jets with $R = 0.4$ (left) and $R = 0.6$ (right).

matched to ROI words at L1 or trigger jets at L2 in order to determine whether the jet should be associated with forward or central trigger. Events are then weighted based on the prescales of the triggers, using a generalized version of the method described in section 1.3.2, in order to account for the larger number of trigger categories.

1.4 Data Correction

after cross section calculations are carried out in section above, data needs to be corrected in order to account for (accpetance) and detector effects. As the jet cross section falls sharply with p_T , the non-zero resolution of the detector causes the measured spectrum to be skewed towards higher p_T from the “true” spectrum. Monte Carlo may be used to compare the p_T of truth jets with reconstructed jets in order to understand the effects of the detector resolution. These effects can then be “unfolded” from the measured spectrum in order to obtain results particle level results from data.

The transfer matrix, A_{ij} , may be derived from Monte Carlo, whereby the entry A_{ij} is the number of truth jets in the j -th p_T bin which are then reconstructed in the i -th p_T bin. The transfer matrix describes the influence that detector effects have on the measured results, and must be derived from simulation. This may be normalized to produce the folding matrix of probabilities, P_{ij} , given by

$$P_{ij} = \frac{A_{ij}}{\sum_{k=0}^{N_b} A_{kj}} \quad (1.12)$$

The spectrum at truth level, t_j , and the reconstructed spectrum r_j , are then related by

$$r_j = \sum_{k=0}^{N_b} P_{jk} t_k. \quad (1.13)$$

An unfolded result, u_j , may then be obtained from the measured data d_j , by solving the

matrix equation

$$d_j = \sum_{k=0}^{N_b} P_{jk} u_k. \quad (1.14)$$

The solution to this may be found by inverting P_{ij} , such that $u_j = \sum_k P_{ik}^{-1} d_k$. However, this is undesirable as it yields large fluctuations in u_j [3]. Transfer matrices tend to have large condition numbers, meaning that the solution is sensitive to slight changes in the input. Small variations in d_j , which may be due to statistical fluctuations, can produce large spurious fluctuations in u_j .

In order to avoid this, the unfolding method must incorporate some form of regularization. Typically equation 1.14 is solved numerically, and regularization may be done, for example, by incorporating the smoothness of u_j into the optimization method [3]. In this analysis, the unfolding is carried out using the Iterative, Dynamically Stabilized (IDS) method [9, 10].

1.4.1 IDS method

In the IDS method, the MC is re-weighted at each iteration such that the reconstructed MC spectrum is brought closer to that of the measured data, with the truth level spectrum and the transfer matrix adjusted accordingly. As the reconstructed MC is brought closer to the measured data spectrum through re-weighting, the truth level spectrum approaches the desired unfolded result. For features present in the measured data which were not simulated by the MC, the difference between data and reconstructed MC is then unfolded.

Regularization is implemented through the use of a regularization function, $f(\Delta x, \sigma, \lambda)$. This function determines how much unfolding should be done at a given iteration based on the difference between data and Monte Carlo, Δx , the uncertainty in this difference, σ , and the regularization parameter λ . The regularization function should take the ratio $\Delta x/\sigma\lambda$ as an argument, and return a value near zero at small values of this argument and approach one at higher values.

As the regularization function depends on the difference between data and Monte Carlo, it is important that the Monte Carlo be normalized appropriately. The discrepancy is defined as

$$\Delta d_k = d_k - \frac{N_{\text{dSmc}}}{N_{\text{MC}}} r_k \quad (1.15)$$

where N_{MC} is the total number of jets in the MC sample and N_{dSmc} is the number of jets in data corresponding to structures which are simulated in the MC. Normalizing the MC in this way enables the unfolding to preserve features in the data, such as new physics signals, which were not simulated in the MC, while correctly scaling the MC in regions where it has a similar shape to the data. The total number of jets in the data sample may be taken as an initial estimate for N_{dSmc} . A better estimate may then be obtained using

$$N'_{\text{dSmc}} = N_{\text{dSmc}} + \sum_{k=1}^{N_B} (1 - f(\Delta d, \sigma, \lambda)) \Delta d_k. \quad (1.16)$$

This procedure may be repeated until the relative change in N_{dSmc} is less than a desired threshold.

1.4.2 Matching Efficiency

In order to obtain the transfer matrix from the Monte Carlo, reconstructed jets need to be matched with a corresponding truth jet. Jets are matched if they lie within a distance $\Delta R < 0.3$ in $y - \phi$ space. The spectrum of matched truth (or reconstructed) jets may be obtained by projecting the transfer matrix on the appropriate axis. The matched spectrum may then be compared to the unmatched spectrum to obtain the matching efficiency for truth jets and for reconstructed jets. As the transfer matrix is derived using only matched pairs of jets, the effect of the matching efficiency should be taken into consideration. Prior to the unfolding the measured data is multiplied by the matching efficiency for reconstructed jets, and after unfolding the result is divided by the

matching efficiency for truth jets.

An initial unfolding may then be performed on the data, yielding the result

$$u_j = t_j \cdot \frac{N_{\text{dSmc}}}{N_{\text{MC}}} + [1 - f(\Delta d_k, \sigma d_k, \lambda_M)] \cdot \Delta d_j + \sum_{k=1}^{N_B} f(\Delta d_k, \sigma d_k, \lambda_M) \cdot \tilde{P}_{kj} \cdot \Delta d_k \quad (1.17)$$

where the unfolding matrix \tilde{P}_{ij} is obtained from the transfer matrix:

$$\tilde{P}_{ij} = \frac{A_{ij}}{\sum_{k=0}^{N_b} A_{ik}} \quad (1.18)$$

The regularization functions determine how much unfolding is done in a given bin, based on the regularization parameter λ and significance of Δd_k . Only the difference between measured data and reconstructed MC is unfolded: In bins where Δd_k is small the unfolded result is dominated by the truth level MC spectrum. In those bins where the discrepancy is significant, the regularization parameter λ determines how much unfolding is carried out. As λ increases, less unfolding is performed on the data.

Once an initial estimate of u_j has been obtained, this may be used to re-weight the MC at truth level, and thus improve the transfer matrix. The updated transfer matrix is given by

$$A'_{ij} = A_{ij} + f(|\Delta u_j|, \sigma u_j, \lambda_M) \cdot \frac{N_{\text{MC}}}{N_{\text{dSmc}}} \cdot P_{ij} \cdot \Delta u_j, \quad (1.19)$$

where

$$\Delta u_j = u_j - \frac{N_{\text{dSmc}}}{N_{\text{MC}}} \cdot t_j. \quad (1.20)$$

The new transfer matrix can then be used to update the folding and unfolding matrices, P_{ij} and \tilde{P}_{ij} .

The procedure may then be carried out iteratively in the following sequence

- Update the normalization factor N_{dSmc} , according to 1.16.
- Perform the unfolding.

- Use the unfolded data to update the transfer matrix, and the folding/unfolding matrices.

1.5 Treatment of uncertainties

1.5.1 Statistical uncertainties

The statistical uncertainties on the final (unfolded) cross section measurement are obtained through “toy” Monte Carlos. For each toy, Poisson fluctuations are applied to the measured data spectrum and to the transfer matrix. The fluctuated data is then unfolded using the fluctuated transfer matrix. The unfolded spectra from the toys are then used to compute a covariance matrix, from which the statistical uncertainty on the unfolded data is taken.

1.5.2 Systematic Uncertainties

The JES uncertainty is the dominant source of systematic uncertainty. Its effect on the final cross section result is found by propagating each of the JES components listed in section 1.2.4 through the unfolding procedure. For each component, the measured data spectrum is shifted up and down in p_T by one standard deviation. The shifted spectra are then unfolded, and the uncertainty associated with that component is taken as the difference between the shifted and unshifted spectra, after unfolding.

The effect of the jet energy resolution (JER) uncertainty is also propagated through the unfolding. A modified Monte Carlo sample is produced by smearing the p_T of the reconstructed jets in the nominal MC sample. The reconstructed jets have their p_T smeared by a factor Δ , where Δ is a random variable with standard deviation σ_Δ and a mean of one. The standard deviation of Δ is chosen to satisfy

$$\sigma_\Delta^2 + \sigma_{\text{nom}}^2 = (\sigma_{\text{nom}} + \sigma_{\text{JER}})^2, \quad (1.21)$$

where σ_{nom} is the nominal JER and σ_{JER} is its uncertainty. The additional smearing is thus applied in such a way as to increase the effective JER by one standard deviation. The data is then unfolded using the modified transfer matrix, and the difference between this result and that unfolded with the nominal transfer matrix is taken as the systematic uncertainty.

The matching between truth and reconstructed jets is another source of systematic uncertainty. As mentioned earlier, the transfer matrix is constructed by matching truth and reconstructed jets within $\Delta R < 0.3$ in rapidity and azimuthal angle. Transfer matrices are also constructed by matching jets within $\Delta R < 0.4$ and $\Delta R < 0.2$. The unfolding is carried out using these matrices, and the largest difference between either result and that obtained from the nominal transfer matrix is taken as the uncertainty.

Shape variations between the MC spectrum and the data will also introduce systematic uncertainty through the unfolding procedure. To estimate this effect, the truth level MC spectrum is reweighted in such a way to improve the agreement between the reconstructed MC and the measured data. The reweighted MC is then unfolded using the original transfer matrix (i.e., the one used to unfold the data). The difference between the result of this and the reweighted truth level MC is taken as a systematic uncertainty, as it reflects the effect of the MC shape on the unfolded spectrum.

The efficiency with which jets are reconstructed has also been considered. This is equivalent to the matching efficiency in the Monte Carlo, and may be estimated in the data by matching reconstructed jets to track jets defined using information from the inner detector. These two efficiencies may be compared in order to estimate the degree to which the jet reconstruction efficiency is mis-modeled by the simulation. The difference is taken as a systematic uncertainty, and is less than 1% for jets with p_T above 30GeV.

plot of relative systematic for certain rapidity regions is shown in figure ??, showing the contributions from various components.

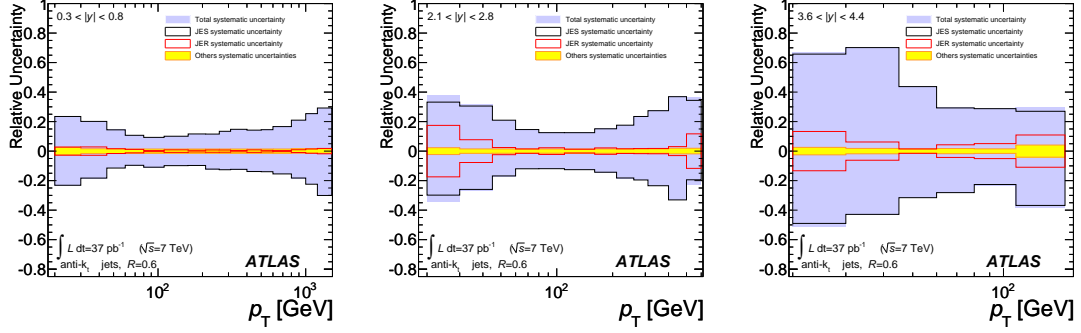


Figure 1.13: Systematic uncertainty on the inclusive jet cross section for anti- k_t jets with $R = 0.6$

1.6 Results and discussion

1.6.1 Inclusive Jet Cross Section

results for the inclusive jet cross section are plotted in figure 1.14 for $R = 0.4$ jets and figure 1.15 for $R = 0.6$. Theoretical predictions are obtained using NLOJET++ with the CT10 pdf set, with the normalization and factorization scales set to the p_T of the leading jet, $\mu_r = \mu_f = p_T^{\max}$. The agreement between data and theory is generally good.

In figures 1.16 ($R = 0.4$) and 1.17 ($R = 0.6$), the ratio of the data and theory results is plotted. The CT10 pdf set is used as a baseline, though results obtained using other pdf sets (MSTW2008, NNPDF 2.1, and HERAPDF 1.5) are also shown. All pdf sets tend to predict a higher cross section than that seen in data, with the discrepancy worse at high p_T or high rapidity. MSTW2008 tends to follow the data better in these regions. In all cases, the differences between data and theory are similar in magnitude to the systematic uncertainties from theory and data.

Figures 1.18 and 1.19 show the same ratios, but also include theoretical results obtained from POWHEG, with the matrix element calculated using the CT10 PDF set and the parton showering carried out using either PYTHIA (with either the AUET2B or Perugia2011 tune) or HERWIG (AUET tune). The POWHEG results vary significantly depending on the method used for parton showering. These variations are on the order

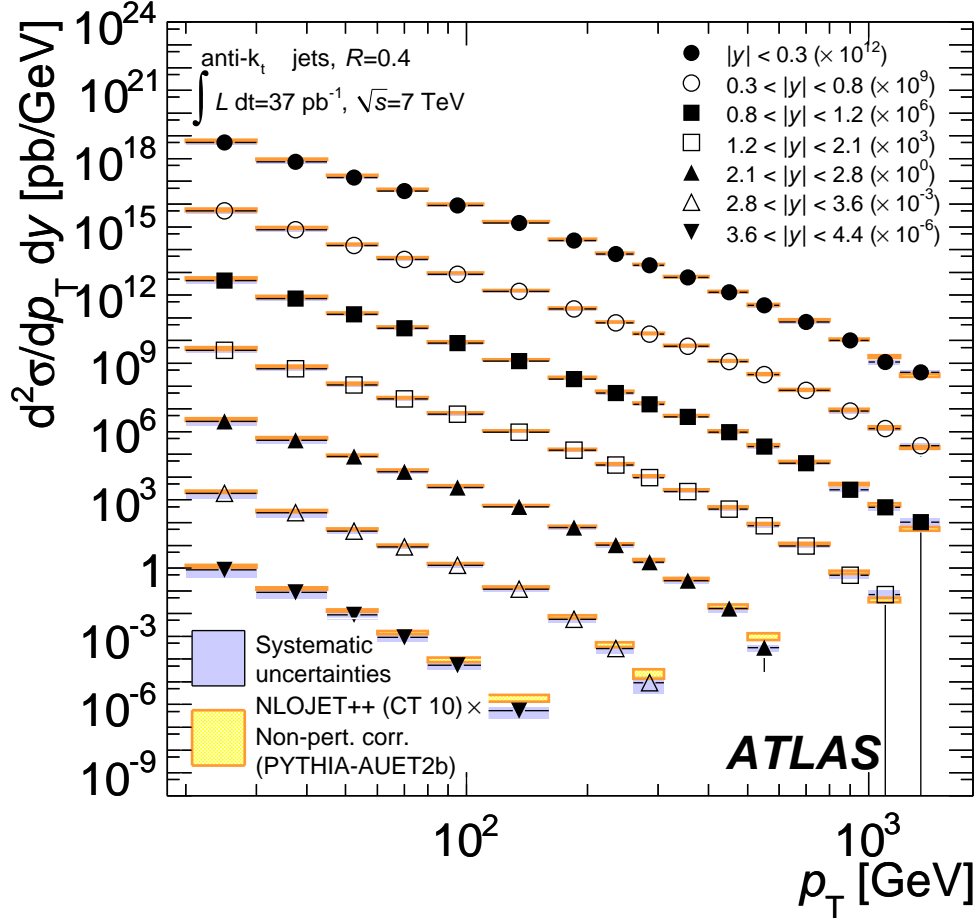


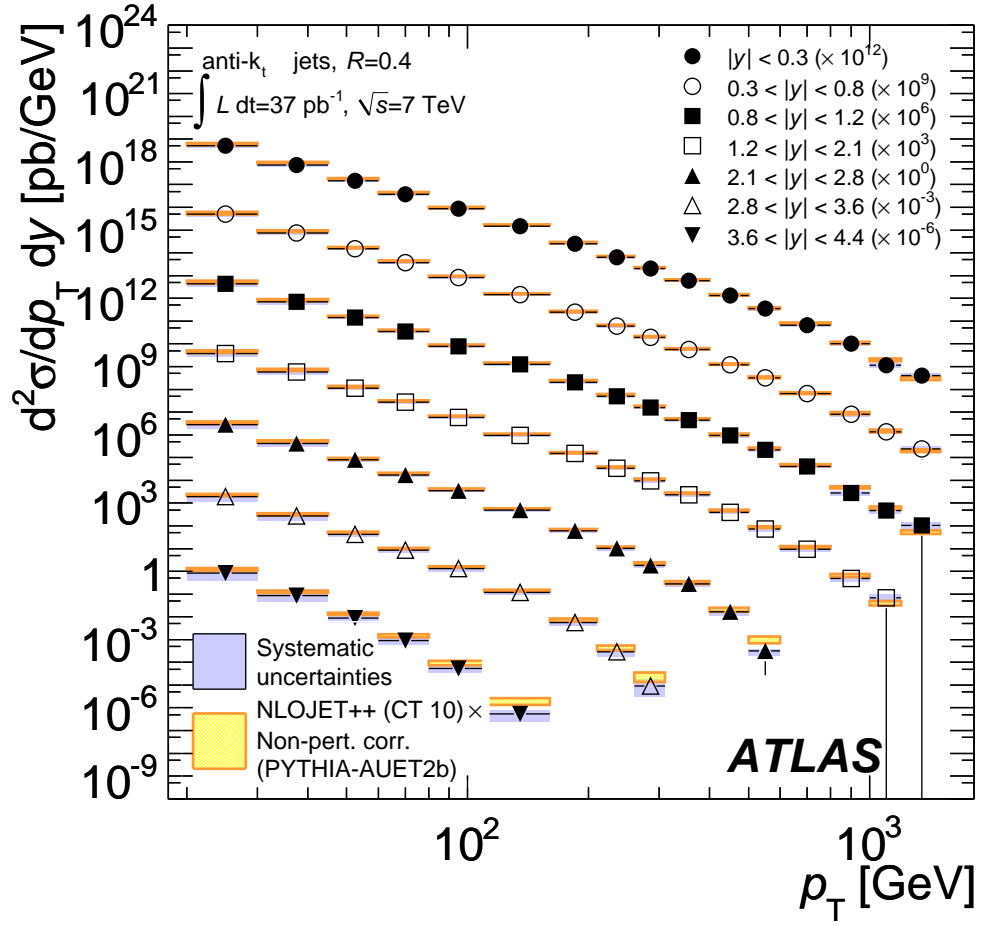
Figure 1.14: Inclusive jet cross section for ant-kt jets with $R = 0.4$.

of 30%, whereas the theoretical uncertainty due to non-perturbative effects is less than 10% for jets with $R = 0.4$.

1.6.2 Dijet Mass Spectrum

The dijet mass spectrum, after unfolding, is plotted in figures 1.20 for jets with $R = 0.4$, and 1.21 for $R = 0.6$. The theoretical results are again obtained using NLOJET++ with the CT10 pdf set and PYTHIA for the parton showering. In this case, the scales

$$\mu_r = \mu_f = \frac{m_{12}}{2 \cosh(0.7y^*)} \quad (1.22)$$

Figure 1.15: Inclusive jet cross section for ant- k_t jets with $R = 0.6$.

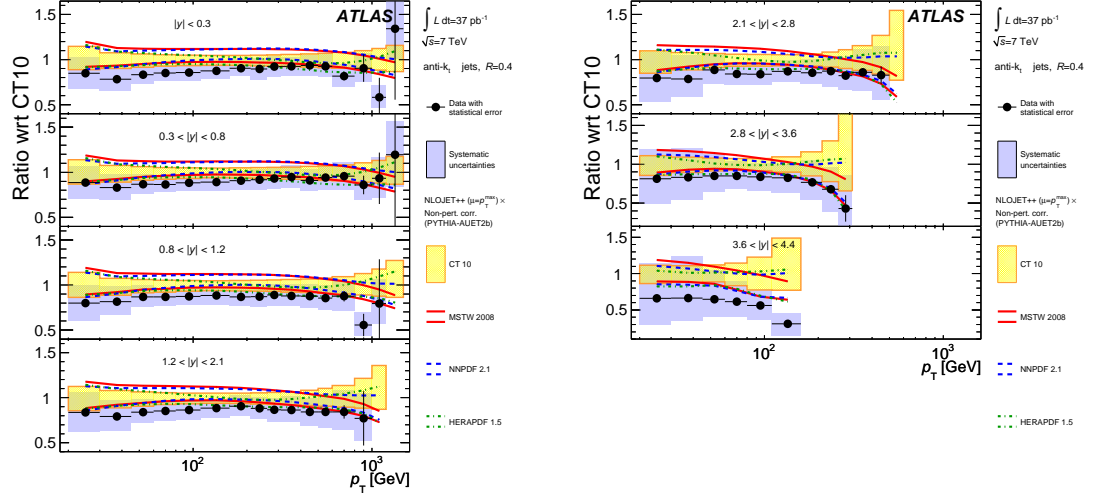


Figure 1.16: Ratio of the inclusive jet cross section to the CTEQ10 prediction, for different rapidity bins using anti- k_t jets with $R = 0.4$. Other pdfs also shown.

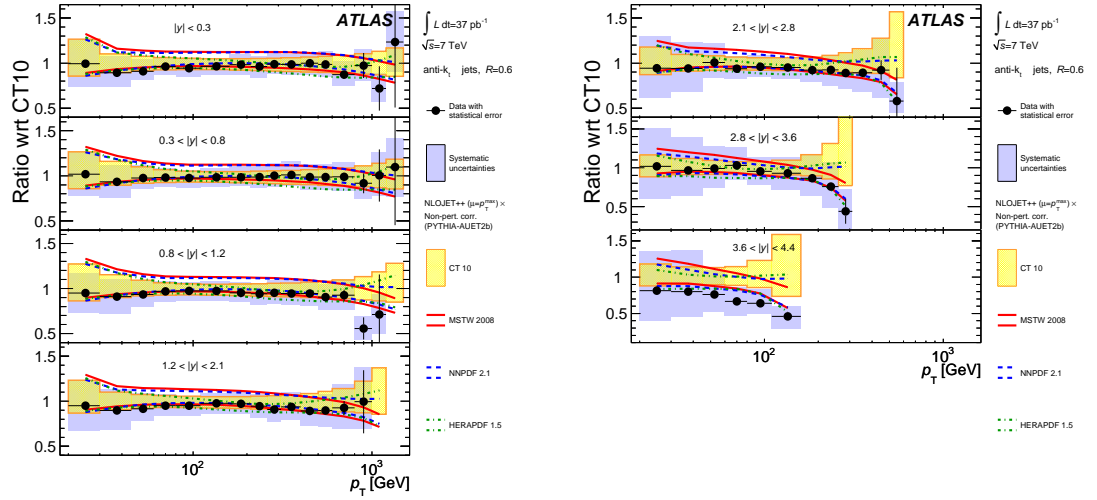


Figure 1.17: Ratio of the inclusive jet cross section to the CTEQ10 prediction, for different rapidity bins using anti- k_t jets with $R = 0.6$. Other pdfs also shown.

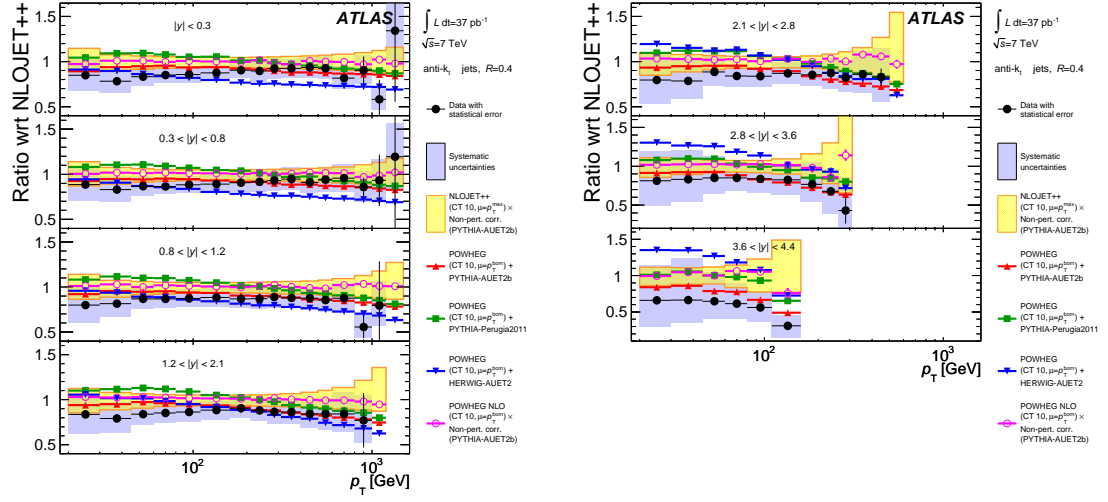


Figure 1.18: Ratio of the inclusive jet cross section to the CTEQ10 prediction, for different rapidity bins using anti- k_t jets with $R = 0.4$. Theoretical predictions made using POWHEG are also shown.

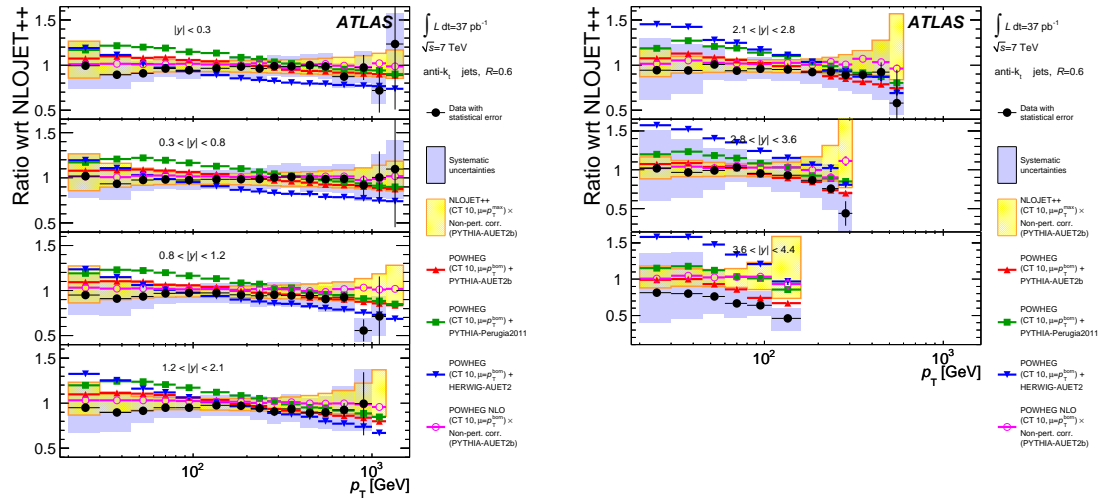
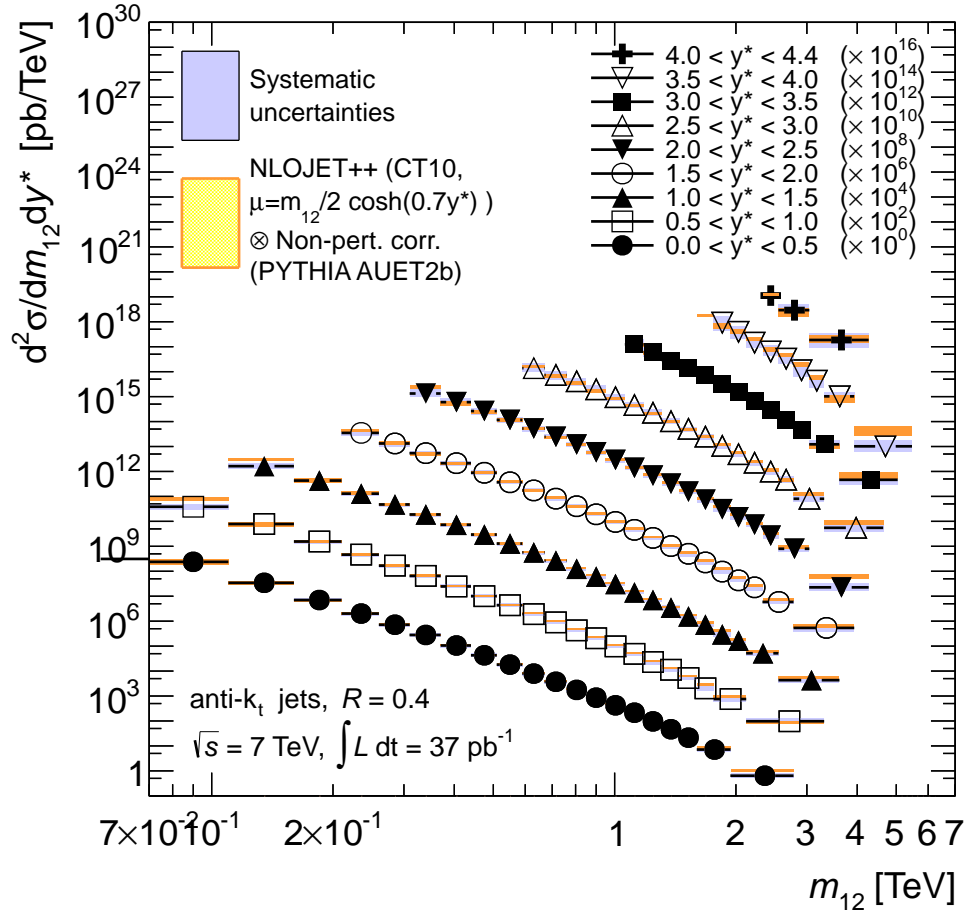


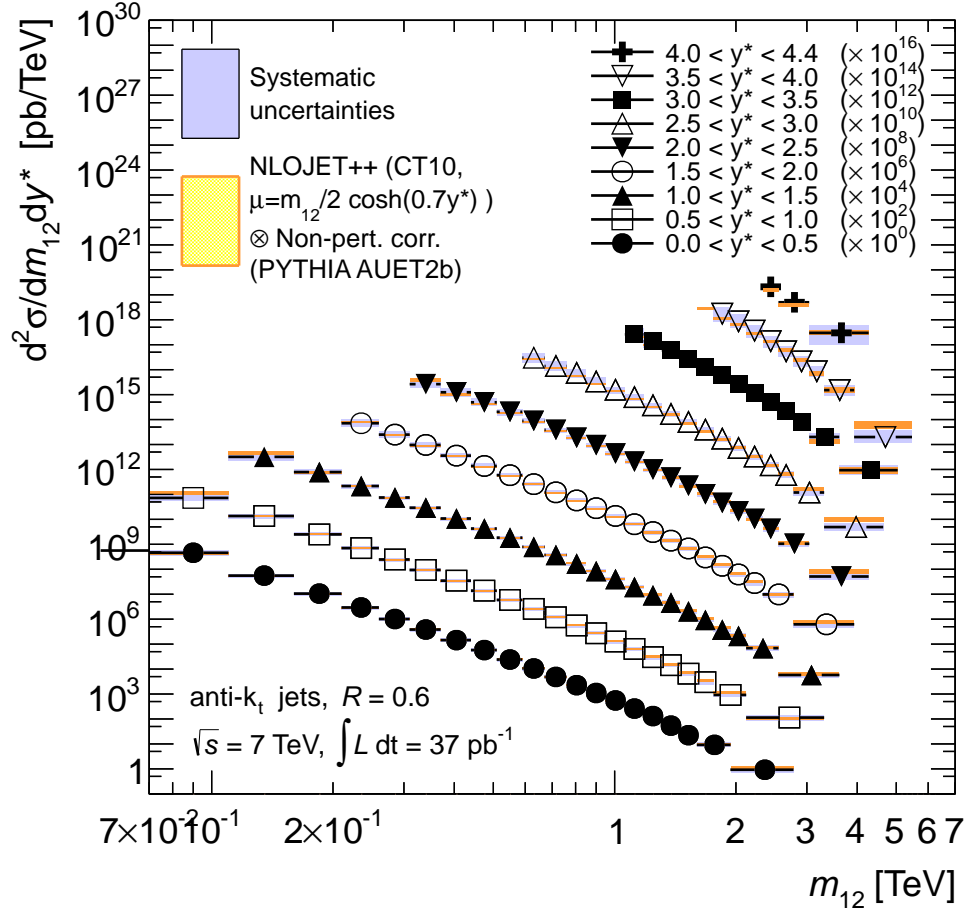
Figure 1.19: Ratio of the inclusive jet cross section to the CTEQ10 prediction, for different rapidity bins using anti- k_t jets with $R = 0.6$. Theoretical predictions made using POWHEG are also shown.

were chosen for normalization and factorization, as this choice was found to be more stable at high values of y^* .

Theoretical results from different pdf sets are plotted in figures 1.22 and 1.23, shown as a ratio to the result obtained using CT10. Data and theory can be seen to agree well at small separations. At larger values of y^* , fluctuations appear in the data/theory ratio. The data spectrum is smooth in this region, and fluctuations are due to lack of statistics in the theoretical predictions.

Results obtained from POWHEG are plotted in figures 1.24 and 1.25. POWHEG did not exhibit the instabilities seen at high separations in NLOJET++, and so the scales $\mu_r = \mu_f = p_T^{\max}$ are again used with POWHEG. The best agreement between theory and data is seen when using the AUET2b tune, with Perugia2011 and AUET yielding cross sections higher than that seen in data.

Figure 1.20: Dijet mass spectrum for ant-kt jets with $R = 0.4$.

Figure 1.21: Dijet mass spectrum for ant-kt jets with $R = 0.6$.

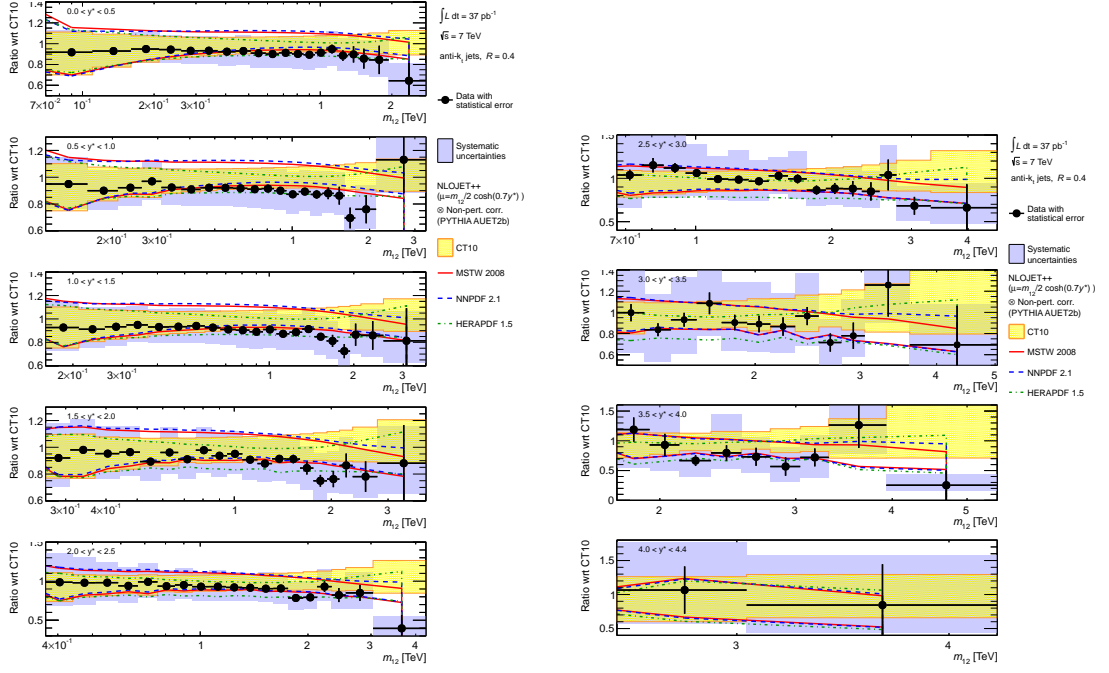


Figure 1.22: Ratio of measured dijet mass spectrum to theoretical prediction obtained using POWHEG, for anti- k_t jets with $R = 0.4$.

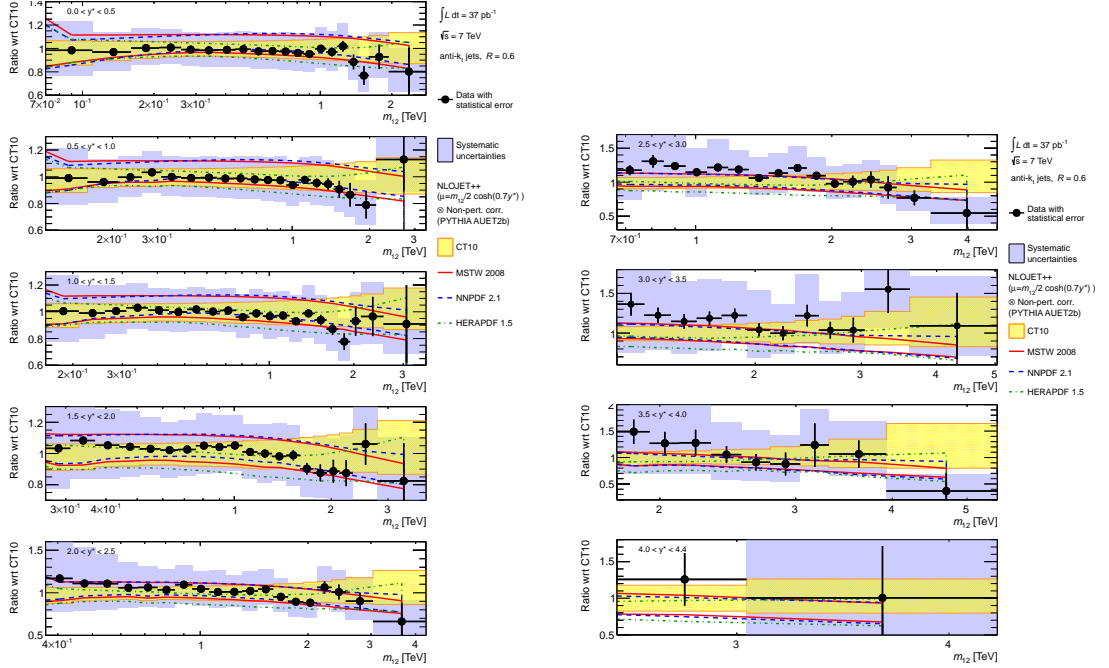


Figure 1.23: Ratio of measured dijet mass spectrum to theoretical prediction obtained using POWHEG, for anti- k_t jets with $R = 0.6$.

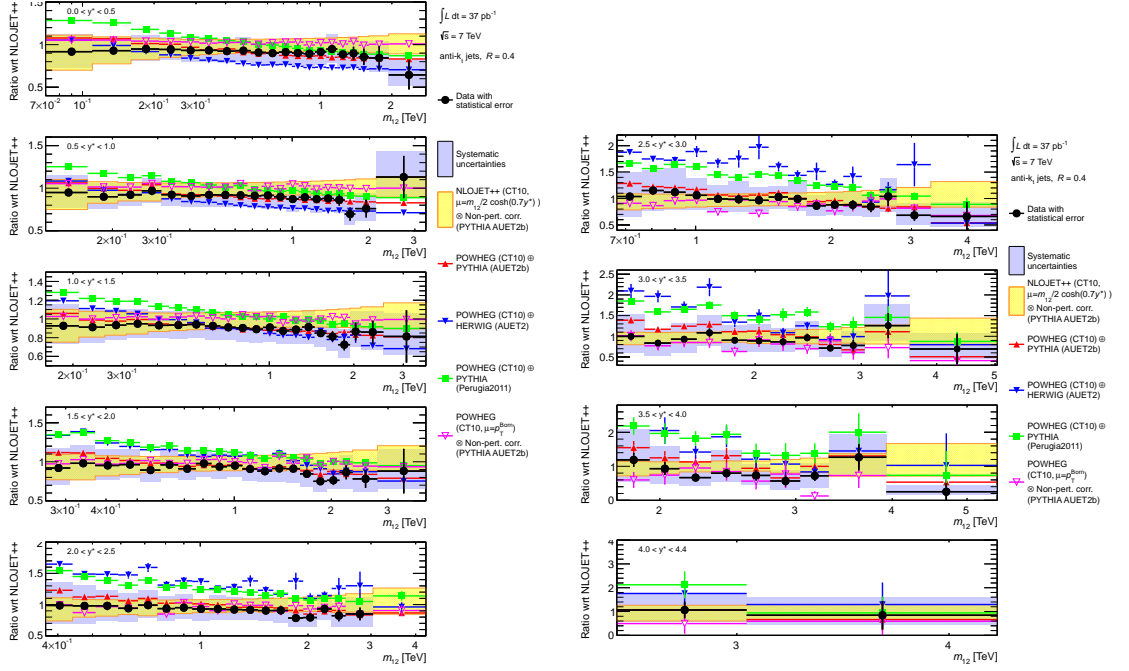


Figure 1.24: Ratio of measured dijet mass spectrum to theoretical prediction obtained using POWHEG, for anti- k_t jets with $R = 0.4$.

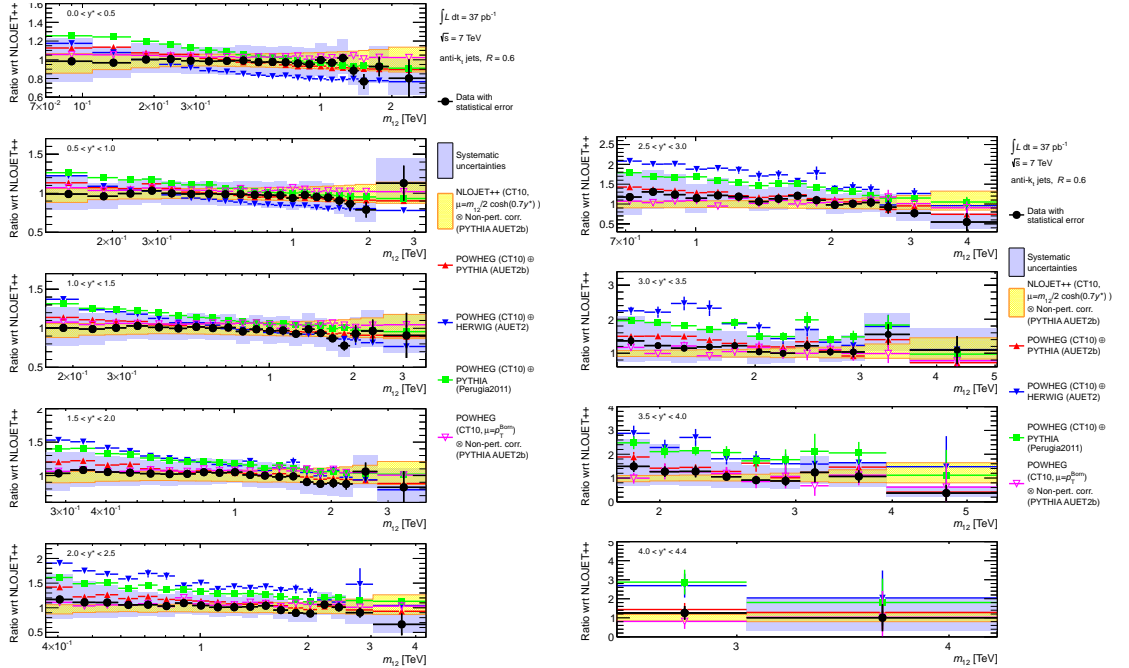


Figure 1.25: Ratio of measured dijet mass spectrum to theoretical prediction obtained using POWHEG, for anti- k_t jets with $R = 0.6$.

Bibliography

- [1] Atlas calorimeter response to single isolated hadrons and estimation of the calorimeter jet scale uncertainty. Technical Report ATLAS-CONF-2011-028, CERN, Geneva, Mar 2011.
- [2] In-situ pseudorapidity intercalibration for evaluation of jet energy scale uncertainty using dijet events in proton-proton collisions at $\sqrt{s}=7$ tev. Technical Report ATLAS-CONF-2011-014, CERN, Geneva, Mar 2011.
- [3] Volker Blobel. An Unfolding method for high-energy physics experiments. pages 258–267, 2002.
- [4] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The antik_t jet clustering algorithm. *JHEP*, 04:063, 2008.
- [5] ATLAS Collaboration. Measurement of inclusive jet and dijet cross sections in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS detector. 2010.
- [6] ATLAS Collaboration. Measurement of jet production in 7 tev proton-proton collisions with the atlas detector. Technical Report ATLAS-COM-CONF-2010-052, CERN, Geneva, Jun 2010. See back up note for full author list.
- [7] E Khramov, N Rusakovich, T Carli, A Henriques, V Gangiobbe, Z Liang, C Santoni, and M Simonyan. Study of the response of the hadronic barrel calorimeter in the

atlas combined test-beam to pions of energies from 20 to 350 gev for beam impact points from 0.2 to 0.65. Technical Report ATL-TILECAL-PUB-2009-007. ATL-COM-TILECAL-2009-006, CERN, Geneva, Apr 2009.

- [8] W Lampl, S Laplace, D Lelas, P Loch, H Ma, S Menke, S Rajagopalan, D Rousseau, S Snyder, and G Unal. Calorimeter clustering algorithms: Description and performance. (ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003), Apr 2008.
- [9] Bogdan Malaescu. An Iterative, dynamically stabilized method of data unfolding. 2009.
- [10] Bogdan Malaescu. An Iterative, Dynamically Stabilized(IDS) Method of Data Unfolding. 2011.